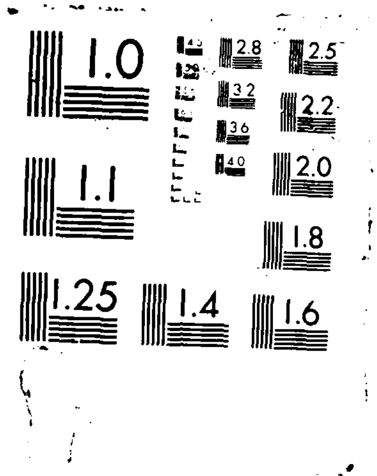
AD-A190 670 RELATIONSHIP OF ENCODING SPEED AND MEMORY TESTS TO FLIGHT TRAINING PERFORMANCE(U) AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TX T R CARRETTA MAR 88 1/1

UNCLASSIFIED AFHRL-TP-87-49 F/G 5/8 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

AFHRL-TP-87-49

DTIC FILE COPY

(2)

AIR FORCE

HUMAN RESOURCES

AD-A190 670

RELATIONSHIP OF ENCODING SPEED AND MEMORY TESTS TO FLIGHT TRAINING PERFORMANCE

Thomas R. Carretta

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

March 1988

Interim Paper for Period September 1983 - December 1986

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

88 3

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-87-49	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b OFFICE SYMBOL (If applicable) AFHRL/MOEA	7a NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b OFFICE SYMBOL (If applicable) HQ AFHRL	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO 62703 · PROJECT NO. 7719 · TASK NO 18 · WORK UNIT ACCESSION NO 45

11. TITLE (Include Security Classification)
Relationship of Encoding Speed and Memory Tests to Flight Training Performance

12. PERSONAL AUTHOR(S)
Carretta, T.R.

13a. TYPE OF REPORT	13b TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM Sep 83 TO Dec 86	March 1988	26

16. SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	
05	08		
05	09		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The demands on the cognitive/perceptual abilities of military pilots have increased steadily as aircraft have become more sophisticated. The ability to encode and classify signals and to retrieve information from short-term memory are two of the several cognitive/perceptual abilities that have been linked to flying performance. Two tests, Encoding Speed (encoding and classification ability) and Immediate/Delayed Memory (short-term memory retrieval), were administered to 2,219 United States Air Force pilot candidates prior to entry into Undergraduate Pilot Training (UPT). Performance on the Encoding Speed test was related to successful completion of UPT, in-flight performance measures, and advanced training assignment. However, scores on the Immediate/Delayed Memory test were not related to training performance. Pilot candidates who made quick and accurate responses on the Encoding Speed test were more likely to perform well on in-flight performance measures and be recommended for post-UPT training in a fast-jet (Fighter-Attack-Reconnaissance) aircraft. Implications for pilot selection and classification are discussed.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b TELEPHONE (Include Area Code) (512) 536-3877 · 22c OFFICE SYMBOL AFHRL/TSR

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

AFHRL Technical Paper 87-49 March 1988

RELATIONSHIP OF ENCODING SPEED AND MEMORY TESTS TO FLIGHT TRAINING PERFORMANCE

Thomas R. Carretta

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

William R. Ercoline, Lt Col, USAF
Chief, Cognitive Skills Assessment Branch

This publication is primarily a working paper. It is published solely to document work performed.

SUMMARY

A variety of cognitive/perceptual abilities have been linked to flying performance. Two tests, Encoding Speed (encoding and classification ability) and Immediate/Delayed Memory (short-term memory retrieval), were administered to 2,219 United States Air Force pilot candidates prior to entry into Undergraduate Pilot Training (UPT). Although both tests were reliable, only performance on the Encoding Speed test was shown to be related to flight training performance. Pilot candidates who made quick and accurate responses on the Encoding Speed test were more likely to perform well on in-flight performance measures and be recommended for additional training in a fast-jet (Fighter-Attack-Reconnaissance) aircraft upon completion of UPT. Implications for pilot selection and early classification (for Specialized Undergraduate Pilot Training) are discussed.

PREFACE

This work was completed under Work Unit 77191845 in support of a Request for Personnel Research (RPR 78-11, Selection for Pilot Training) submitted by training program managers.

This paper is intended to serve as an interim report regarding two of the cognitive/perceptual tests of the Basic Attributes Tests (BAT) battery.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RELATIONSHIP OF ENCODING SPEED AND MEMORY TESTS
TO FLIGHT TRAINING PERFORMANCE

I. INTRODUCTION

As aircraft have become increasingly sophisticated with modern technological advances, many functions previously performed by the pilot have been automated. These advances, however, have increased the demands placed on the pilot's cognitive abilities. Crucial to pilot performance are: the speed and accuracy with which stimuli are perceived, encoded, stored, transformed, and compared; the speed with which memory is searched and accessed; and the speed with which decisions are made.

Most of the research in the area of cognitive factors related to pilot performance has focused on pretraining techniques to improve comprehension and integration of information needed to fly the aircraft (Crosby, 1977; Gerlach, 1974; Jensen & Benel, 1977). When tests of cognitive abilities have been used for pilot selection rather than for training, the emphasis has been on selective attention (Gopher & Kahneman, 1971) and time-sharing ability (North & Gopher, 1976). Recent test batteries designed to assess the performance of aerospace ground equipment crews (McLaurin, 1973), Air Force navigator trainees (Hunter, 1975), and advanced simulator trainees in Undergraduate Pilot Training (UPT) (Pew, Rollins, Adams, & Gray, 1977) have included several cognitive tasks involving memory, spatial visualization, verbal reasoning, and other abilities.

A review of these and other studies helped lead to the development of a computer-administered test battery, the Basic Attributes Tests (BAT), designed to improve the selection and classification of United States Air Force pilot and navigator trainees. The original BAT battery consisted of 15 tests; it measured psychomotor skills, as well as a variety of cognitive/ perceptual abilities and personality/attitudinal characteristics believed to be related to pilot and navigator performance (see Carretta, 1987a, for a more complete description of the BAT battery).

Several of the cognitive abilities tests in the BAT battery have been evaluated in terms of their ability to predict various flight performance measures and final training outcome (see Table 1). These include Digit Memory (information input efficiency), Decision-Making Speed (choice reaction time), Item Recognition (short-term memory storage, search, and comparison operations), Mental Rotation (spatial transformation), and Time-Sharing (higher-order tracking ability, and learning rate and time-sharing ability as a function of differential task load). Of particular interest was the potential of these cognitive tests, which rely heavily on paper-and-pencil measures, to increase the validity of current selection procedures used by the U.S. Air Force to select pilot candidates. Although the experimental cognitive tests investigated did not add significantly to the prediction of graduation or elimination from UPT, they did demonstrate significant relationships with several other performance measures including recommendations for fighter assignments after training (Carretta, 1987a, 1987b, 1987c). Pilot candidates who made quick, consistent, and accurate responses on these tests were more likely to perform better during flight training and receive a post-training assignment for a fighter aircraft.

This paper examines the predictive utility of two remaining cognitive tests in the BAT battery; namely, Encoding Speed (classification) and Immediate/Delayed Memory (short-term memory retrieval). The general hypotheses guiding this investigation were that individual differences in performance on the tests would predict UPT performance and also that use of the tests in combination with the Air Force Officer Qualifying Test (AFOQT) would add significantly to the validity of predictions concerning flight training success (Rogers, Roach, & Wegner, 1986). In particular, it was expected that subjects with quicker reaction times and greater response

accuracy would be more likely to succeed in training, and that these differences would be reflected in better flight performance scores (check flight scores), which have a broader range than the dichotomous final training outcome measure (UPT pass/fail). The fact that the pass/fail rate is unevenly distributed (80% pass versus 20% fail) also makes that criterion less sensitive.

Table 1. Cognitive/Perceptual Tests in the BAT Battery

Test	Reference	Cognitive/Perceptual ability
Digit Memory	Hunter, 1975	Perceptual speed, information input efficiency
Decision-Making Speed	Fleishman and Hempel, 1955	Choice reaction time
Item Recognition	Sternberg, 1966	Short-term memory storage, search and comparison
Mental Rotation	Shepard and Metzler, 1971	Spatial transformation
Time-Sharing	North and Gopher, 1976	Selective attention
Encoding Speed	Posner and Mitchell, 1967	Classification speed
Immediate/ Delayed Memory	Hunter, 1975	Memory retrieval

If the scores from the two cognitive tests, taken together with AFOQT scores, failed to demonstrate a stronger relationship with training performance outcomes than the AFOQT alone, there would be no reason to go to the effort to modify or replace the current test system. Conversely, if the cognitive tests were found to add to the validity of the current test procedure, this could be interpreted as evidence that the cognitive tests measure unique abilities unrelated to those associated with the AFOQT.

In addition to predicting successful completion of pilot training, the Air Force is interested in classifying pilots, as early as possible, for advanced training. Currently, UPT lasts 49 weeks and includes about 175 hours of flying time. Based on an evaluation by an Advanced Training Recommendation Board (ATRB) in the 42nd week of UPT, pilot candidates are recommended for one of two advanced training tracks: Fighter-Attack-Reconnaissance (FAR) aircraft or Tanker-Transport-Bomber (TTB) aircraft. In general, the students who perform best during UPT are selected for fast-jet training (i.e., FAR). Therefore, it was hypothesized that the FAR-recommended pilots would demonstrate quicker, more consistent, and more accurate performance on the cognitive/perceptual tests than would the TTB-recommended pilots. The demonstration of such a relationship would enable the Air Force to identify potential FAR and TTB candidates early in UPT and thus result in more efficient and cost-effective training.

II. METHOD

Subjects

The subjects in this study were 2,219 Air Force officer candidates targeted for UPT. They were tested on the Encoding Speed and Immediate/Delayed Memory tests prior to their entry into UPT. Only subjects who also had scores on the AFOQT were included in the regression analyses

that predicted performance on the UPT final outcome measure. Pilot training performance measures were available for only a portion of these subjects, as many of them had not yet completed UPT.

Procedure

Each subject was tested on the AFOQT prior to entry into pilot training. This test provided five composite scores based on several subtests: Verbal, Quantitative, Academic Aptitude (verbal and quantitative combined), Navigator-Technical, and Pilot. Only the Navigator-Technical and Pilot composite scores have been used in the operational selection of candidates for UPT (USAF, 1983). The predictive utility of the AFOQT-Pilot composite score served as a baseline to judge the usefulness of the Encoding Speed and Immediate/Delayed Memory tests for improving pilot selection and classification for specialized training, as this composite has demonstrated a consistent relationship with pilot training performance (Arth, 1986).

The Encoding Speed and Immediate/Delayed Memory tests were included in the BAT battery. After a test administrator initiated the system, the test session was self-paced by the subject. The test session lasted about 3 1/2 hours and included scheduled breaks between tests to avoid problems with mental and physical fatigue.

AFOQT-Pilot Composite

The AFOQT-Pilot composite score is based on performance on 8 of the 16 AFOQT subtests. These subtests assess a variety of skills including verbal reasoning, mechanical and instrument comprehension, scale and table reading, spatial transformation, and general aviation knowledge.

Encoding Speed

According to Posner and Mitchell (1967), performance on this task reflects verbal classification ability at several levels of cognitive processing.

In this task, subjects were presented simultaneously with a pair of letters and required to make a same-different judgment about the letter pair. This judgment was based on one of three decision rules: Physical Identity (AA versus Aa), Name Identity (AA versus AE), or Category Identity (vowels versus consonants - AE versus AH). The latency of the encoding judgment provides a measure of the speed of the encoding process. Latency differences among the three types of judgments indicate the speed of recoding. For instance, the reaction time for the Name Identity judgments minus reaction time for Physical Identity judgments is assumed to indicate the speed with which physical stimuli may be recoded to the level at which their name may be accessed.

The Encoding Speed test consisted of three 32-item blocks of trials, which required about 15 minutes to complete. Response time and accuracy were recorded on each trial.

Immediate/Delayed Memory

This test was designed to assess continuous short-term memory storage and retrieval operations.

In this task, the subject was presented with a sequence of digits and required to respond by indicating the digit that occurred either one or two digits previously. There were two parts to each of these subtasks. In the first part, the digits were presented for 1/2 second, followed by a 2-second inter-stimulus interval. In the second part, the inter-stimulus interval was 5 seconds. Thus, for both the one-back and two-back subtasks, part one dealt with "immediate" memory, and part two with "delayed" memory.

There were 25 trials in each level of subtask (one- versus two-back) by length of latency (2 versus 5 seconds) condition, resulting in 100 trials. As with Encoding Speed, response time and accuracy were recorded on each trial. The Immediate/Delayed Memory test required about 25 minutes for completion.

UPT Performance Criteria

UPT final outcome was assigned at the completion of UPT and was recorded as a dichotomous variable (fail = 0 and pass = 1). Subjects who satisfactorily completed UPT received a recommendation for follow-on training in either a fast-jet (FAR) or a slower aircraft (TTB) by an Advanced Training Recommendation Board (ATRB) consisting of T-38 Instructor Pilots (TTB = 0 and FAR = 1). Generally, FAR aircraft are considered to be more demanding than TTB aircraft.

UPT final outcome and ATRB recommendation were determined, in part, by a subject's performance on six check flights during UPT. A check flight involved an in-flight performance evaluation by an Instructor Pilot. The first three check flights took place in a T-37, low-performance jet trainer. Three later flights took place in a T-38, higher-performance supersonic jet trainer. The T-37 check flights included: midphase contact, a subject's first check flight; contact, in which the subject's ability to perform maneuvers and aerobatics by visual cues from outside the plane was evaluated; and instrument, in which the subject was required to perform maneuvers by reference to the displays on cockpit instruments. The T-38 check flights, in addition to contact and instrument, included evaluation of the subject's ability to fly in formation with other aircraft. Each subject received an overall check flight grade (1-unsatisfactory, 2-fair, 3-good, or 4-excellent) and a percentage score (based on performance of certain maneuvers within the flight) for all check flights completed during training. Table 2 provides a summary of the number of subjects who had scores on the two BAT tests and the UPT performance criteria.

Table 2. Number of Subjects

Tests	Test only	UPT	ATRB	Check flights
Encoding Speed	2,219	930	625	184
Immediate/Delayed Memory	867	567	390	132
Both Tests	867	545	376	118

III. RESULTS AND DISCUSSION

Encoding Speed

Descriptive Measures

Table 3 summarizes the average response times and percent correct scores for the same and different judgments for each of the three types of encoding decisions. Accuracy of response was high for all parts of this test (85.7% to 95.3% correct). This was considered encouraging. In keeping with common practice regarding tasks of this nature, calculation of average response times was based only on those trials with correct responses.

Table 3. **Encoding Speed: Average Response Time and Percent Correct**

Condition	Number of trials	Response time (ms) Mean	SD	% Correct
Physical Identity				
same	16	662.9	196.1	95.3
different	16	758.8	200.9	94.5
		[710.9]		[94.9]
Name Identity				
same	16	695.0	162.7	92.3
different	16	741.1	176.4	94.5
		[718.1]		[93.4]
Category Identity				
same	16	925.1	248.8	85.7
different	16	940.5	236.1	92.7
		[932.8]		[89.2]

Note. N = 2,219.

Average response time was slightly lower for the 32 Physical Identity trials (mean = 710.9 ms.) than for the Name Identity trials (mean = 718.1 ms.), but was considerably longer for the Category Identity trials (mean = 932.8 ms.). Although the order of the response time means was consistent with Posner and Mitchell's (1967) "depth of processing" model, the difference in average response times between Physical Identity and Name Identity trials was considerably less than that reported by Posner and Mitchell (1967) (10 ms. instead of 50 ms.). Contrary to their depth of processing model, the average response time for correct "different" judgment trials was somewhat longer for Physical Identity trials (mean = 758.8 ms.) than for Name Identity trials (mean = 741.4 ms.). These results are depicted in Figure 1.

For the present investigation, Cronbach's reliability coefficient indicated that the 96 items (alpha = .711) and their response times (alpha = .958) were reliable.

Factor Structure

As the pattern of response times for these subjects did not fit the depth of processing model, differences in response times between types of judgment were not calculated. Average response time for correct responses, percent correct, and a response time by percent correct interaction term were calculated for the Physical, Name, and Category Identity trials separately and used in the factor analysis. These variables were chosen to reflect three important aspects of judgment: speed, accuracy, and speed-accuracy tradeoffs for differing levels of judgment.

As indicated in Table 4, the strongest inter-item correlations occurred between variables of the same type. The response time measures were related strongly to each other ($.587 \leq r \leq .738$) but only weakly to the accuracy scores ($.200 \leq r \leq .215$) and interaction terms ($-.175 \leq r \leq -.017$). The accuracy scores were related moderately to each other ($.286 \leq r \leq .457$), but were not related to the interaction terms ($-.076 \leq r \leq .043$). The interaction terms were not related to each other ($-.028 \leq r \leq .099$).

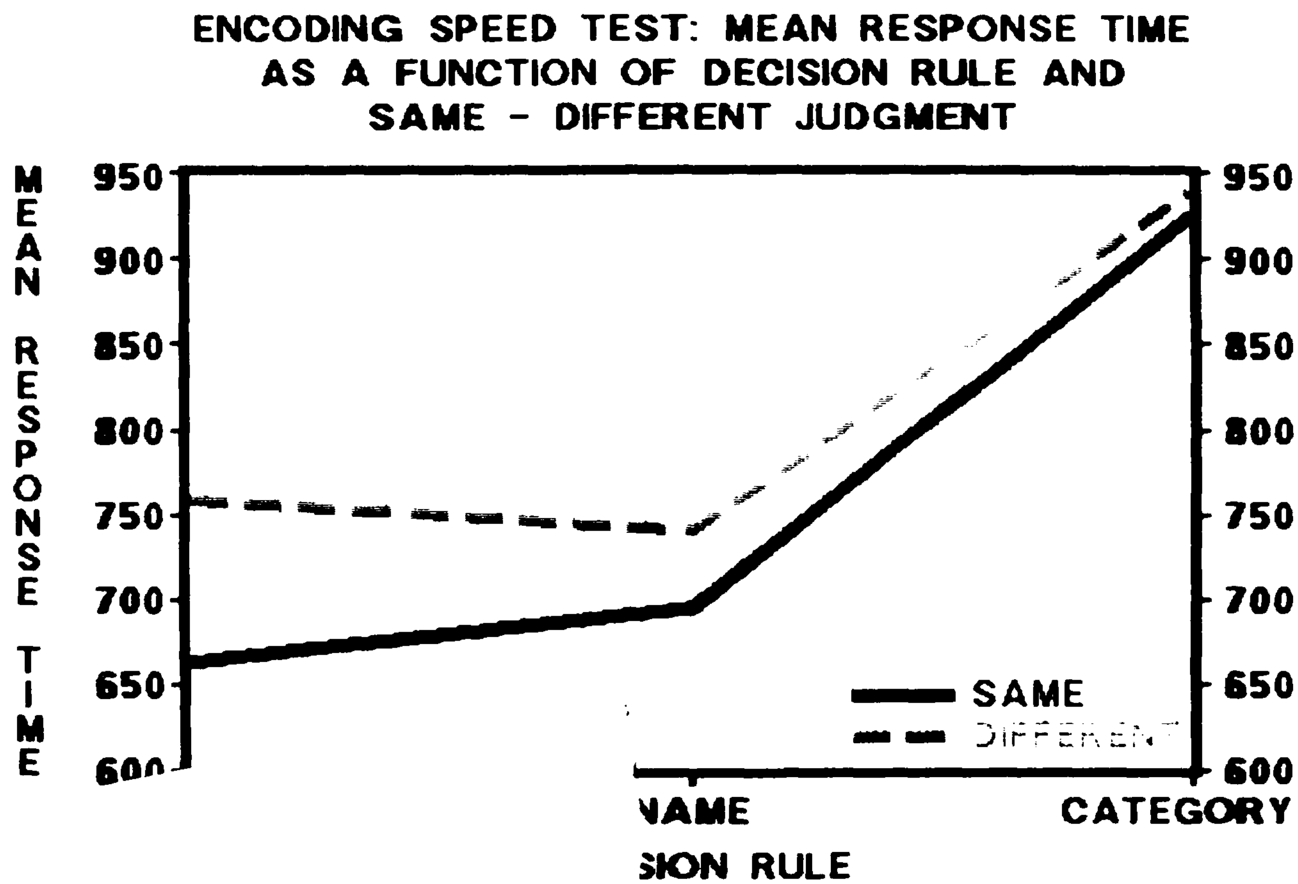


Figure 1. Encoding Speed: Average Response Time by Type of Judgment.

Table 4. Encoding Speed: Inter-Item Correlation Matrix

Variable	1	2	3	4	5	6	7	8	9
1. Average RT - Physical	1.000								
2. Average RT - Name	.633	1.000							
3. Average RT - Category	.587	.738	1.000						
4. % Correct - Physical	.208	.198	.194	1.000					
5. % Correct - Name	.156	.215	.179	.296	1.000				
6. % Correct - Category	.106	.298	.200	.286	.457	1.000			
7. RT by % - Physical	-.029	.032	.003	-.076	-.069	-.085	1.000		
8. RT by % - Name	.064	-.017	.060	-.025	.043	-.101	.099	1.000	
9. RT by % - Category	-.124	-.160	-.175	-.015	-.078	-.031	-.028	-.001	1.000

Note. Variable labels refer to the average response time, percent correct, and average response time by percent correct interaction term for the Physical, Name, and Category Identity conditions, respectively.
N = 2,219.

The factor analysis yielded three factors that accounted for 57.7% of the total item variance. After rotation, the principal factor, which consisted of the three average response times, accounted for 65.4% of the explained variance. The three accuracy scores loaded on Factor 2, and the three interaction terms loaded primarily on Factor 3. Results of the factor analysis are shown in Table 5.

Table 5. Encoding Speed: Summary of Factor Analysis

Variable	Communality	Factor 1	Factor 2	Factor 3
1. Average RT - Physical	.509	.702	.104	.074
2. Average RT - Name	.816	.879	.187	-.088
3. Average RT - Category	.688	.816	.139	.050
4. % Correct - Physical	.187	.174	.392	-.061
5. % Correct - Name	.539	.101	.724	.065
6. % Correct - Category	.472	.145	.631	-.229
7. RT by % - Physical	.037	.025	-.109	.157
8. RT by % - Name	.274	.034	.010	.522
9. RT by % - Category	.036	-.183	-.038	-.029

Factor	Eigenvalue	% of Explained variance	Cumulative %
1	2.33	65.4	65.4
2	.89	25.0	90.4
3	.34	9.6	100.0

Note. N = 2,219.

These results suggested that the most conceptually important measure provided by this test was general response latency. The accuracy of response and speed by accuracy tradeoff variables were less important; however, they did make unique contributions in terms of explaining performance on this test. Based on these results, three variables were retained for the flight training regression analyses: overall response time, percent correct, and a response time by percent correct interaction term.

Inferential Measures

To evaluate the usefulness of the cognitive tests for predicting flight training performance, it was necessary to determine the extent to which the current selection instrument (i.e., AFOQT-Pilot composite score) was related to performance. The AFOQT-Pilot composite score demonstrated a modest but significant relationship only with UPT final outcome (r = .090, $p \leq$ .05) and was related marginally to advanced training assignment (r = .099, $p \leq$ .10). The AFOQT-Pilot composite score was not related to check flight performance. Results of the regression analyses are summarized in Table 6.

Table 6. AFOQT-Pilot Composite Score: Summary of UPT Outcome Regression Analyses

Outcome measure	N	Outcome measure Mean	Outcome measure SD	AFOQT-Pilot Mean	AFOQT-Pilot SD	r
UPT (pass/fail)	545	0.79	0.41	71.97	17.84	.090*
ATRB (TTB/FAR)	376	0.60	0.49	73.46	17.41	.099
T-37 midphase grade	118	2.52	1.18	69.75	19.46	.115
T-37 contact grade	117	2.97	0.94	70.05	19.32	-.001
T-37 instrument grade	115	2.93	1.01	70.35	19.11	.151
T-38 contact grade	105	2.51	1.16	70.83	19.54	-.001
T-38 instrument grade	103	2.87	1.11	70.85	19.58	.092
T-38 formation grade	101	2.84	1.04	70.98	19.70	.085
T-37 midphase percentage	118	95.03	8.76	69.75	19.46	.021
T-37 contact percentage	117	91.13	5.38	70.05	19.32	.100
T-37 instrument percentage	115	91.63	7.56	70.35	19.11	.070
T-38 contact percentage	105	91.19	5.75	70.83	19.54	.064
T-38 instrument percentage	103	91.72	11.48	70.85	10.58	.154
T-38 formation percentage	101	92.70	6.74	70.98	19.70	.067

*$p \leq$ .05.

As can be seen in Table 7, the Encoding Speed model was not related significantly to UPT final outcome (R = .101, n.s.) but was related to ATRB rating (R = .165, $p \leq$ .025). It should be noted that average response latency was related differentially to final training outcome and advanced training recommendation. Subjects who made quick responses were less likely to complete UPT. However, of those who successfully completed UPT, those who made quick responses were more likely to be recommended for advanced training in fast jets (FAR aircraft). One explanation for these results is that during the early part of training (T-37 phase), cautious actions (slower responses) are considered desirable. As a result, some "reckless" quick-responders are eliminated. During the later phases of training when faster aircraft are used (T-38 phase), the ability to respond quickly becomes an important asset. This explanation is supported by the pattern of correlations between average response time and T-38 check flight performance scores. Average response time is negatively correlated with performance for each of the six T-38 check flight grades and percentage scores. Performance during the T-38 phase, in turn, has a direct impact on the advanced training assignment. The Encoding Speed model was related significantly to performance on two of the three T-37 check flight grades: contact (R = .274, $p \leq$ .05) and instrument (R = .262, $p \leq$ .05). It was not related to T-38 performance.

Table 7. Encoding Speed: Summary of UPT Outcome Regression Analyses

Outcome measure	N	Correlation with outcome			Multiple R			R^2 Change
		Average RT	% Correct	RT by %	Encoding Speed	AFOQT-Pilot	Combined model	
UPT (pass/fail)	545	.081*	-.025	-.001	.101	.090*	.156*	.016*
ATRB (TTB/FAR)	376	-.151*	-.011	-.109	.165*	.099	.176*	.021*
T-37 midphase grade	118	-.066	.073	-.112	.133	.115	.161	-----
T-37 contact grade	117	-.060	.079	-.273*	.274*	-.001	.275(.06)	.076*
T-37 instrument grade	115	.135*	.094	-.182*	.262*	.151	.310*	.073*
T-38 contact grade	105	-.045	-.145	.068	.146	-.001	.153	-----
T-38 instrument grade	103	-.135	-.142	-.053	.188	.092	.229	-----
T-38 formation grade	101	-.150	-.159	.049	.191	.085	.227	-----
T-37 midphase percentage	118	-.159	-.064	-.154	.217	.021	.220	-----
T-37 contact percentage	117	.078	.079	-.133	.173	.100	.198	-----
T-37 instrument percentage	115	.009	-.104	-.095	.202	.070	.237	-----
T-38 contact percentage	105	-.018	-.108	.041	.109	.064	.153	-----
T-38 instrument percentage	103	-.043	-.088	-.015	.100	.154	.218	-----
T-38 formation percentage	101	-.035	-.104	.072	.109	.067	.150	-----

Note. The change in R square between the AFOQT-Pilot composite and the combined model was tested only when the combined model was judged to be significant. *Similarly, the components of the Encoding Speed model* were tested only when the Encoding Speed model and/or the combined model was judged to be significant. The value in parentheses indicates the significance level of the combined model when it was marginally significant.

*$p \leq .05$.

A combined model that used the AFOQT-Pilot composite score and the three Encoding Speed scores was related significantly to UPT final outcome ($R = .156$, $p \leq .01$), ATRB rating ($R = .176$, $p \leq .025$), and the T-37 instrument flight grade ($R = .310$, $p \leq .05$) and was related marginally to the T-37 contact flight grade ($R = .275$, $p \leq .10$). In each of these instances, the combined model significantly improved prediction of performance beyond the level of prediction provided by the AFOQT-Pilot composite score alone. These results suggested that the Encoding Speed test measured some flight training performance-related ability not captured by the AFOQT-Pilot composite score. The combined model multiple correlation increases are summarized in the last column of Table 7.

Immediate/Delayed Memory

Descriptive Measures

Percent correct and average response time for correct responses for each of the four parts are summarized in Table 8. Accuracy of response was high for the first, second, and fourth parts of this test (averaging 91.5% correct), but was rather low for part three (68.7% correct). This may have occurred because at part three the task changed, requiring subjects to remember a digit that was presented two digits back (delayed memory) rather than one digit back. Some of the subjects may have been confused by the instructions for part 3 or may have found the changed task more difficult. Whatever the reason for the decrement in performance, the subjects recovered during part 4 (delayed memory - 5-second delay). Despite this problem, responses on this test were very reliable (alpha = .938).

Table 8. **Immediate/Delayed Memory:**
Average Response Time and Percent Correct

Condition	Number of trials	Response Time (ms.) Mean	SD	% Correct
One Digit Back				
2-second delay	25	496.2	257.0	89.1
5-second delay	25	457.1	217.9	94.6
Two Digits Back				
2-second delay	25	551.3	308.2	68.7
5-second delay	25	493.6	252.8	90.8

Note. $N = 867$.

Average response time for correct responses was fairly consistent across the four parts of this test (496 ms., 457 ms., 551 ms., and 494 ms., respectively) and was very reliable over the 100 trials (alpha = .943).

Results from the accuracy and response time measures suggested that there was an accuracy by response time interaction due primarily to the part 3 trials. Response time was longer and accuracy was lower during part 3.

Factor Structure

The most conceptually important measures from this test were average response time and percent correct for each of the four parts, as these measures provided information regarding

individual differences in the speed and accuracy of immediate and delayed memory operations. Speed by accuracy interaction terms for the four parts were included in the factor analysis, as speed/accuracy tradeoffs were also considered important.

The inter-item correlation matrix, presented in Table 9, indicated that the four average response time measures were related positively to each other ($.287 \leq r \leq .553$) and negatively to their respective percent correct scores ($-.459 \leq r \leq -.285$). These results provided additional evidence of a speed/accuracy tradeoff, as subjects with quicker response times tended to be less accurate. The four percent correct scores were correlated moderately with each other ($.144 \leq r \leq .397$). The interaction terms were not related to each other ($-.027 \leq r \leq .144$), but were related to the response time and percent correct scores that contributed to the interaction term.

The factor analysis yielded five factors which accounted for 73.7% of the total item variance. Factors 1, 2, 4, and 5 consisted of the average response time, percent correct, and the response time by percent correct interaction term for each of the four stimulus conditions (1- versus 2-back by 2- versus 5-second delay). Factor 3 can be thought of as a general "response latency" factor, as the four response times loaded on it. A summary of the final factor solution is provided in Table 10.

These results suggested that although there was a common general response latency component for the four parts of this test (Factor 3), the task demands were unique, in part, for each of the four parts of this test (as indicated by Factors 1, 2, 4, and 5). As a result, all 12 variables were retained for use in the prediction of flight training performance.

Inferential Measures

The Immediate/Delayed Memory model, with all 12 variables from the factor analysis, demonstrated poor predictive utility against all of the performance criteria. The model was not related to UPT pass/fail outcome ($R = .172$, n.s.), advanced training recommendation (TTB/FAR) ($R = .213$, n.s.), or check flight performance. The zero-order correlations between the 12 variables in the Immediate/Delayed Memory model and the various performance criteria were not consistently in the same direction. This lack of stability was not surprising as very few of these zero-order correlations were significant at the .05 level of probability. A summary of these regression analyses is provided in Table 11.

As can be seen from Table 11, in general, neither the AFOQT-Pilot composite score nor the Immediate/Delayed Memory model was related closely to flight training performance. Subjects with higher AFOQT-Pilot composite scores were more likely to successfully complete UPT ($r = .090$, $p \leq .05$), whereas the Immediate/Delayed Memory model was not related significantly to any of the training performance measures.

A combined model was tested to determine whether the AFOQT-Pilot composite score and the Immediate/Delayed Memory model, taken together, could improve prediction of flight training performance. As can be seen from the last two columns of Table 11, the combined model was only marginally related to one of the performance measures: UPT pass/fail ($R = .193$, $p \leq .10$). The Immediate/Delayed Memory model did not add significantly to the predictive utility provided by the AFOQT-Pilot composite alone.

Full Model

When considered separately, neither the AFOQT-Pilot composite score nor the BAT tests demonstrated a close, consistent relationship with all three performance criteria. The

Table 9. **Immediate/Delayed Memory: Inter-Item Correlation Matrix**

Variable (digit, delay)	1	2	3	4	5	6	7	8	9	10	11	12
1. Avg RT1 (1, 2 sec)	1.000											
2. Avg RT2 (1, 5 sec)	.446	1.000										
3. Avg RT3 (2, 2 sec)	.338	.287	1.000									
4. Avg RT4 (2, 5 sec)	.345	.382	.553	1.000								
5. % Correct1 (1, 2 sec)	-.459	-.301	-.094	-.067	1.000							
6. % Correct2 (1, 5 sec)	-.165	-.387	-.109	-.171	.397	1.000						
7. % Correct3 (2, 2 sec)	-.199	-.060	-.410	-.162	.195	.146	1.000					
8. % Correct4 (2, 5 sec)	-.166	-.093	-.182	-.285	.144	.176	.321	1.000				
9. RT x %1 (1, 2 sec)	-.491	-.087	-.026	-.037	.579	.127	.095	.059	1.000			
10. RT x %2 (1, 5 sec)	-.006	-.628	-.003	.007	.243	.468	.054	.005	.013	1.000		
11. RT x %3 (2, 2 sec)	.051	.117	-.347	-.091	-.016	-.005	.193	.099	.015	-.012	1.000	
12. RT x %4 (2, 5 sec)	-.010	.002	-.099	-.458	.060	.153	.091	.308	.008	-.027	.144	1.000

Note. Variable labels refer to the average response time, percent correct, and average response time by percent correct interaction term for the one digit back - 2-second delay, 1 digit back - 5-second delay, two digits back - 2-second delay and 2 digits back - 5-second delay conditions, respectively.

N = 867.

Table 10. **Immediate/Delayed Memory: Summary of Factor Analysis**

Variable (digit/delay)	Communality	Factor loadings				
		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1. Avg RT1 (1, 2 sec)	.685	-.027	-.372	.162	-.060	-.018
2. Avg RT2 (1, 5 sec)	.923	-.704	.098	.703	.099	.024
3. Avg RT3 (2, 2 sec)	.928	-.016	.124	.261	-.778	-.037
4. Avg RT4 (2, 5 sec)	.752	-.030	.061	.261	-.186	-.557
5. % Correct 1 (1, 2 sec)	.692	.317	.463	.214	.091	.086
6. % Correct 2 (1, 5 sec)	.400	.550	.032	.081	.092	.207
7. % Correct 3 (2, 2 sec)	.303	.082	.041	.071	.502	.121
8. % Correct 4 (2, 5 sec)	.233	.070	.020	.061	.241	.389
9. RT x %1 (1, 2 sec)	.545	-.009	.293	-.008	.010	.012
10. RT x %2 (1, 5 sec)	.771	.876	-.089	.349	-.006	-.065
11. RT x %3 (2, 2 sec)	.212	-.040	.000	.055	.441	.098
12. RT x %4 (2, 5 sec)	.576	.014	.036	.072	.072	.755

Factor	Eigenvalue	% of explained variance	Cumulative %
1	2.84	40.4	40.4
2	1.57	22.3	62.7
3	1.18	16.9	79.6
4	0.78	11.1	90.7
5	0.65	9.3	100.0

Note. $N = 867$.

Table 11. **Immediate/Delayed Memory: Summary of UPT Outcome Regression Analyses**

		Correlation with outcome measure													Multiple R		
Outcome Measure	**N**	**Avg. RT1**	**Avg. RT2**	**Avg. RT3**	**Avg. RT4**	**%Correct 1**	**%Correct 2**	**%Correct 3**	**%Correct 4**	**RT by % 1**	**RT by % 2**	**RT by % 3**	**RT by % 4**	**I/D memory**	**AFOQT Pilot**	**Combined model**	**R^2Change**
UPT (pass/fail)	545	.060	-.017	-.074	-.095	-.018	-.002	.042	.059	-.064	.005	.068	.021	.172	.090*	.193(.09)	.029
ATRB (TTB/FAR)	376	-.030	.014	-.071	-.123	.103	.038	.118	.126	.087	-.040	.066	.001	.213	.099	.227	
T-37 midphase grade	118	.079	.146	.212	.141	-.104	-.037	.000	.069	-.112	-.105	-.008	-.15	.310	.115	.338	
T-37 contact grade	117	.036	.045	.064	-.026	-.056	.020	-.019	.031	-.002	-.005	.042	-.118	.227	-.001	.228	
T-37 inst. grade	115	.071	.028	.045	.004	-.038	-.008	.017	.106	.007	-.011	.045	.041	.199	.151	.253	
T-38 contact grade	105	-.211	-.122	-.019	-.070	.229	.084	.159	.073	-.145	.047	-.101	-.125	.405	-.001	.427	
T-38 inst. grade	103	-.088	.052	-.166	-.068	-.058	-.124	.133	-.144	.013	-.112	-.046	-.014	.357	.092	.375	
T-38 formation grade	101	.142	.130	.107	.081	-.051	-.154	-.042	-.076	.085	-.110	.060	.005	.233	.085	.335	
T-37 midphase perc.	118	.084	.204	.147	.129	-.180	-.149	-.039	-.036	-.008	-.166	.072	.017	.264	.021	.277	
T-37 contact perc.	117	.104	.052	.092	.064	-.037	-.021	.009	.021	.036	.015	.160	.088	.222	.100	.244	
T-37 inst. perc.	115	.072	.122	.039	.015	.068	-.114	-.026	-.023	.159	-.088	.073	.010	.277	.070	.306	
T-38 contact perc.	105	-.204	-.094	-.004	-.073	.135	.088	.200	.004	.073	.003	-.096	.038	.413	.064	.417	
T-38 inst. perc.	103	-.028	.083	-.044	-.078	.030	-.065	.237	-.087	.032	-.059	.008	.012	.358	.154	.329	
T-38 formation perc.	101	.068	.104	-.006	.031	-.036	-.066	.104	.067	.054	-.078	.012	-.072	.246	.067	.266	

Note. The change in R square between the AFOQT-Pilot composite and the combined model was tested only when the combined model was judged to be significant. The value in parentheses indicates the significance level of the combined model when it was marginally significant.

*$p \leq .05$.

AFOQT-Pilot composite score was related to UPT final outcome, but not to advanced training assignment or check flight performance. In contrast, the Encoding Speed model was not related to UPT final outcome, but was related to advanced training assignment and performance on two of the three T-37 check flights. Subjects who made quick responses were more likely to be recommended for advanced training in fast-jet aircraft. Scores from the Immediate/Delayed Memory model were not related at all to pilot training performance.

As the AFOQT-Pilot composite, Encoding Speed model, and Immediate/Delayed Memory model demonstrated different patterns of relationships with the flight training performance measures, it was felt that each of these tests may have been measuring some unique ability. If this was so, prediction of performance might be improved by using measures from more than one source in an integrated model.

Each combined model was evaluated to determine whether scores from the experimental tests (Encoding Speed and Immediate/Delayed Memory) improved prediction of flight training performance beyond the level of prediction provided by the currently used selection measure (AFOQT-Pilot composite score). A combined model that used the AFOQT-Pilot composite along with the Encoding Speed measures significantly improved the prediction of UPT final outcome, advanced training recommendation, and two of the three T-37 check flight grades (see Table 7). The Immediate/Delayed Memory measures used in combination with the AFOQT-Pilot composite, however, did not improve prediction of flight training performance (see Table 11).

A "full model" that used the AFOQT-Pilot composite score along with both the Encoding Speed and Immediate/Delayed memory measures was related significantly only to UPT final outcome (R = .230, p .05). The full model did predict final outcome better than the AFOQT-Pilot composite alone (r = .090) ($F[15,529]$ = 1.67, p .05) but did not differ significantly from a reduced model consisting of the AFOQT-Pilot composite and the Encoding Speed measures only (R = .156) ($F[12,532]$ = 1.34, n.s.). The full model regression analyses are summarized in Table 12.

A comparison of the various combined models suggested that the Encoding Speed model was able to improve flight training performance prediction beyond that provided by the AFOQT-Pilot composite, whereas the Immediate/Delayed Memory model was not. Based on these results, the Immediate/Delayed Memory test should probably not be retained in the BAT battery.

IV. CONCLUSION

Performance measures from both cognitive tests were sufficiently reliable to be used for selection purposes; however, only scores from the Encoding Speed test were related to flight training performance. Those from the Immediate/Delayed Memory test were not.

Encoding Speed scores, when used in combination with a currently used selection instrument (AFOQT-Pilot composite score), improved the prediction of several flight training performance criteria, including successful completion of training, advanced training recommendation, and check flight performance. It should be noted, however, that average response time on the Encoding Speed test was related differentially to final training outcome versus T-38 performance and advanced training recommendation. Slower, more cautious subjects were more likely to complete training successfully. Among those who graduated from UPT, however, those who made quick and accurate responses on the Encoding Speed test were more likely to perform well on the T-38 check flights and receive a post-UPT recommendation for advanced training in a FAR aircraft. The latter relationship is consistent with results from other cognitive tests in the BAT battery (Decision-Making Speed, Item Recognition, Mental Rotation, Time-Sharing; see Carretta, 1987a, 1987b, 1987c). These cognitive tests may be most useful when it is desirable

Table 12. Full Model: Summary of UPT Outcome Regression Analyses

Outcome measure	N	Multiple R: AFOQT Pilot	Multiple R: Encoding Speed	Multiple R: I/D Memory	Multiple R: AFOQT-Pilot + Encoding Speed	Multiple R: AFOQT-Pilot + I/D Memory	Multiple R: Full Model	Compared with AFOQT-Pilot R^2 Change
UPT (pass/fail)	545	.090*	.101	.172	.156*	.193	.230*	.045*
ATRB (TTB/FAR)	376	.099	.165*	.213	.176*	.227	.259 (.06)	.057
T-37 midphase grade	118	.115	.133	.310	.161	.338	.359	----
T-37 contact grade	117	-.001	.274*	.227	.275 (.06)	.228	.357	----
T-37 inst. grade	115	.151	.262*	.199	.310*	.253	.377	----
T-38 contact grade	105	-.001	.146	.405	.153	.427	.449	----
T-38 inst. grade	103	.092	.188	.357	.229	.375	.407	----
T-38 formation grade	101	.085	.191	.233	.227	.335	.409	----
T-37 midphase perc.	118	.021	.217	.264	.220	.277	.363	----
T-37 contact perc.	117	.100	.173	.222	.198	.244	.298	----
T-37 inst. perc.	115	.070	.202	.277	.237	.206	.388	----
T-38 contact perc.	105	.064	.109	.413	.153	.417	.437	----
T-38 inst. perc.	103	.154	.100	.358	.218	.329	.418	----
T-38 formation perc.	101	.067	.109	.246	.150	.266	.319	----

Note. The change in R square between the AFOQT-Pilot composite and the full model was tested only when the full model was judged to be significant. Values in parentheses indicate the significance level of the full model when it was marginally significant.

*$p \le .05$.

to make specialized track assignments early in training or when only fighter-recommended (FAR) or non-fighter-recommended (TTB) pilots are needed (Euro-NATO Joint Jet Pilot Training or Air National Guard units).

The next step in the evaluation of the seven BAT cognitive/perceptual tests will be to determine whether they make unique contributions to the prediction of flight training performance when considered together. If they are redundant, some of them may be removed from the BAT battery and replaced by tests of other abilities.

REFERENCES

Arth, T.O. (1986). Validation of the AFOQT for non-rated officers (AFHRL-TP-85-50, AD-A164 134). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987a). Basic Attributes Test (BAT) system: A preliminary evaluation (AFHRL-TR-87-20, AD-A188 503). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987b). Spatial ability as a predictor of flight training performance (AFHRL-TP-86-70, AD-A183 141). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Carretta, T.R. (1987c). Time-sharing ability as a predictor of flight training performance (AFHRL-TP-86-69, AD-A181 838). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Crosby, J.V. (1977). Cognitive pretraining: An aid in the transition from instrument to composite flying (AFHRL-TR-77-62, AD-A048 816). Williams AFB, AZ: Air Force Human Resources Laboratory.

Fleishman, E.A., & Hempel, N.E. (1955). The relation between abilities and improvement with practice in a visual discrimination reaction task. Journal of Experimental Psychology, 49, 301-312.

Gerlach, V.S. (1974). Cues, feedback and transfer in undergraduate pilot training, phase III. Tempe, AZ: Arizona State University, Department of Education Technology.

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criterion. Perceptual and Motor Skills, 33, 1335-1342.

Hunter, D.R. (1975). Development of an enlisted psychomotor/perceptual test battery (AFHRL-TR-75-60, AD-A020 544). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Jensen, R.S., & Benel, R.A. (1977). Judgment evaluation and instruction in civil pilot training. Champaign, IL: Illinois University at Urbana-Champaign, Savoy Aviation Research Laboratory.

McLaurin, W.A. (1973). Validation of a battery of performance tests for prediction of aerospace ground equipment course grades (AFHRL-TR-73-20, AD-774 586). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

North, R.A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. Human Factors, 18, 1-14.

Pew, R.W., Rollins, A.M., Adams, M.J., & Gray, T.N. (1977). Development of a test battery for selection of subjects for ASPT experiments (Report No. 3585). Cambridge, MA: Bolt, Beranek and Newman, Inc.

Posner, M.I., & Mitchell, R.F. (1967). Chronometric analyses of classification. Psychological Review, 74, 392-409.

Rogers, D.L., Roach, B.W., & Wegner, T.G. (1986). Air Force Officer Qualifying Test Form O: Development and standardization (AFHRL-TR-86-24, AD-A172 037). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dementional objects. Science, 171, 701-703.

Sternberg, S. (1966). High speed scanning in human memory. Science, 153, 652-654.

United States Air Force. (1983, July). Application procedures for Undergraduate Flying Training (UNT) (AF Regulation 51-4). Washington, DC: Department of the Air Force.

END
DATE
FILMED
4-88
DTIC